

광고 글 필터링 모델 적용 및 성능 향상 방안

박래근, 윤혁진, 신의철, 안영진, 정승도*
상명대학교 스마트정보통신공학과

Application of Advertisement Filtering Model and Method for its Performance Improvement

Raegeun Park, Hyeok-Jin Yun, Ui-Cheol Shin, Young-Jin Ahn, Seungdo Jeong*
Department of Smart Information and Telecommunication Engineering, Sangmyung University

요약 최근 기하급수적인 인터넷 데이터의 증가로 딥러닝 등의 많은 분야가 발전하였지만 바이럴 마케팅(viral marketing)과 같은 상업적 목적의 광고가 발견되면서 정보증가의 부작용이 발생하고 있다. 이는 양질의 정보를 공유하고자 하는 인터넷의 본질을 훼손하고 있을 뿐만 아니라 사용자는 양질의 정보를 습득하기 위해 검색시간이 증가하는 문제가 야기된다. 이에 본 연구에서는 광고(Ad: Advertisement, 이하 Ad) 글을 정보 전달의 본질을 흐리는 내용의 글이라 정의하였으며 본 정의에 부합하는 정보로 필터링하는 모델을 제안하였다. 제안하는 모델은 광고 필터링 경로와 광고 필터링 성능 개선경로로 구성되었으며 지속적으로 성능이 개선되도록 설계하였다. 광고 글 필터링을 위해 데이터를 수집하고 KorBERT를 사용하여 문서분류를 학습하였다. 본 모델의 성능을 검증하기 위해 실험을 진행하였으며 5개의 주제를 통합한 데이터에 대한 정확도(Accuracy), 정밀도(Precision)는 각각 89.2%, 84.3%의 결과를 나타냈고 광고의 비정형적 특성을 고려하더라도 높은 성능이 보임을 확인하였다. 본 모델을 통해 바이럴 마케팅으로 구성된 문서에서 광고 문단을 판단하고 필터링하여 사용자에게 양질의 정보를 효과적으로 전달하며 검색하는 과정에서 낭비되는 시간과 피로가 감소할 것으로 기대된다.

Abstract In recent years, due to the exponential increase in internet data, many fields such as deep learning have developed, but side effects generated as commercial advertisements, such as viral marketing, have been discovered. This not only damages the essence of the internet for sharing high-quality information, but also causes problems that increase users' search times to acquire high-quality information. In this study, we define advertisement as "a text that obscures the essence of information transmission" and we propose a model for filtering information according to that definition. The proposed model consists of advertisement filtering and advertisement filtering performance improvement and is designed to continuously improve performance. We collected data for filtering advertisements and learned document classification using KorBERT. Experiments were conducted to verify the performance of this model. For data combining five topics, accuracy and precision were 89.2% and 84.3%, respectively. High performance was confirmed, even if atypical characteristics of advertisements are considered. This approach is expected to reduce wasted time and fatigue in searching for information, because our model effectively delivers high-quality information to users through a process of determining and filtering advertisement paragraphs.

Keywords : Advertisement, Unstructured Data, Filtering, Natural Language Processing, KorBERT

이 논문은 한국전자통신연구원에서 공개한 한국어 언어모델(KorBERT)를 사용함(No.2013-2-00131, 휴먼 지식증강 서비스를 위한 지능진화형 WiseQA 플랫폼 기술 개발)

*Corresponding Author : Seungdo Jeong(Sangmyung Univ.)

email: sdjeong@smu.ac.kr

Received July 15, 2020

Accepted November 6, 2020

Revised August 12, 2020

Published November 30, 2020

1. 서론

현재 인터넷이 급속도로 발전하면서 많은 사용자가 스마트폰, PC 등을 통한 SNS, 클라우드, 포털사이트를 일상적으로 사용하고 있다. IDC 보고서에 따르면 인터넷 사용자는 2019년 약 41억 명으로 2005년 대비 약 30억 명 증가했으며[1] 데이터 또한 2018년 기준 33ZB, 2025년 175ZB로 예측되고 있다[2]. 인터넷상의 정보량이 늘어나면서 사용자들은 어떠한 정보를 찾기 위한 수단으로 서적 참조와 같은 고전적 방법이 아닌 인터넷 검색이 일상이 되었다. 인터넷의 발전으로 딥러닝 기술을 위한 데이터 셋의 구성이 용이해져 비약적인 발전이 있었으며 현재도 진행 중이다. 정보의 증가는 기술의 발전과 사용자의 편의에 도움이 되는 양상을 보였으나 다른 부작용이 발생하기도 한다.

인터넷을 통해 사용자들은 나이, 성별, 사회적 신분과 상관없이 자유롭게 개인의 경험, 사진, 생각을 공유할 수 있다. 이에 따르는 부작용으로 인터넷의 특성을 악용한 광고 글이 생성되고 있다. 광고 글은 사용자 자신의 경험을 자발적으로 올려 다른 사용자에게 도움이 되는 요소가 아닌 상업적 목적의 행위로서 인터넷이 가지는 본질을 흐리며 일반 사용자의 정보 습득을 위한 검색 시간의 증가를 야기한다. 이와 관련하여 국내에서 광고 등 원하지 않는 내용을 차단하는 기능을 담은 소프트웨어에 대한 법적 분쟁이 발생하였으며 “포털사이트에 접속한 개별 사용자들이 거기에서 제공되는 광고 등 콘텐츠를 본래의 형태와 내용 그대로 열람해야 할 의무가 없다”는 판결이 나왔다[3]. 이와 같이 인터넷 상에서의 광고는 단순한 홍보를 넘어서 양질의 정보를 습득하고자 하는 사용자와 홍보를 목적으로 하는 기업간의 갈등으로 문제가 심화되고 있다[4].

의견을 자유롭게 공유하고 사용자에게 양질의 정보를 제공하고자 하는 인터넷의 본질이 훼손되는 상황이 지속적으로 발생하면서 사용자의 정보검색 목적에 부합하지 않는 정보를 필터링하는 해결방안이 필요해지고 있다. 무분별한 광고를 필터링하기 위한 기술로서 대표적으로 Adblock이 개발되어 사용되고 있지만 광고로 등록된 이미지를 차단하는 형태의 서비스이다. 사람들이 인터넷에서 접하게 되는 정보의 대부분은 글로 작성되어 있어 Adblock으로 광고 전체를 필터링하기에는 한계가 있다. 본 연구에서는 인터넷에서 점차적으로 정보의 질을 해치는 교묘한 광고문제를 해결하기 위한 광고 글 필터링 모델을 제안한다. 2장에서는 기존 광고 필터링과 관련된 연

구 및 정의를 간략하게 살펴보고 3장에서는 광고글 필터링을 위해 본 논문에서 제안하는 모델에 대해서 제시한다. 4장에서는 제안 모델에 대하여 구체적으로 기술하고 5장에서 실험 및 분석을 통한 성능평가 결과를 제시하고 6장에서 결론을 맺는다.

2. 광고 필터링 관련 연구

광고란 “명시적 비명시적 광고 주체가 목표 고객에게 브랜드 자산을 구축하기 위해 직간접 매체를 활용해 내용을 전달하는 마케팅 커뮤니케이션 활동”이며, “명시적 광고 주체가 광고 목표를 달성하기 위해 유료 혹은 무료로 매체를 활용해 내용을 전달하는 브랜드 커뮤니케이션 활동”이자, “광고 주체가 광고 대상의 우호적인 태도와 행동을 유발하기 위해 직간접 매체를 활용해 아이디어와 제품 및 서비스 내용을 전달하는 단계별 설득 커뮤니케이션 활동”이다[5]. 이러한 광고를 통해 제품을 홍보하는 등 큰 마케팅 효과를 얻을 수 있으나 최근 무분별한 바이럴 마케팅으로 인한 문제점이 발생하고 있다.

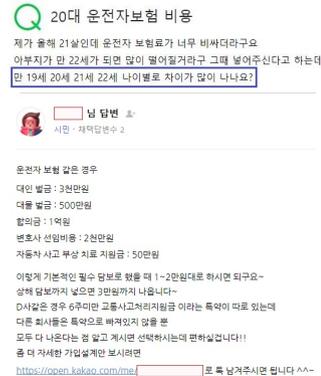


Fig. 1. Example of excessive viral marketing

바이럴 마케팅이란 어떤 기업이나 회사의 제품을 소비자의 힘을 빌려 바이러스처럼 확산되는 방식의 마케팅이라는 의미이며 우리나라의 경우 포털사이트를 통한 연관 검색어, 실시간 검색어, 블로그, 지식iN 포스팅에서 주로 볼 수 있고 최근에는 페이스북, 트위터 등 SNS 매체에서도 많이 찾아볼 수 있다[6]. 이러한 바이럴 마케팅은 주제의 논점을 흐리고 정보의 질을 떨어트리는 등 사용자의 정보검색 활동에 피해를 줄 수 있다.

그림 1은 네이버 지식iN에서 운전자보험을 검색하여 나온 결과 중 하나이다[7]. 질문자는 나이별 보험 비용 차

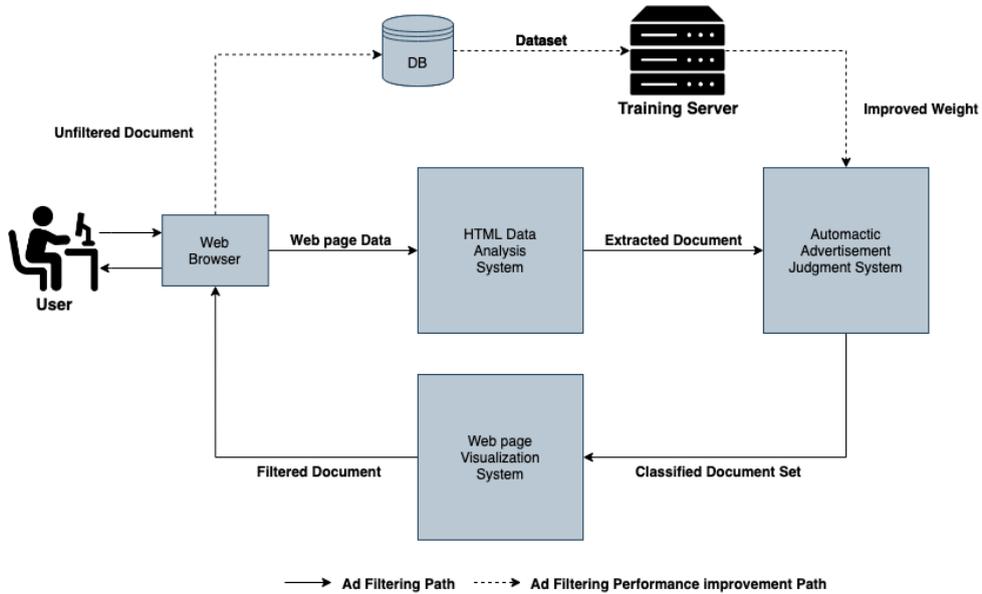


Fig. 2. The Proposed Advertisement Filtering Model

이에 대해 질문했지만 답변자는 질문의 의도에 벗어나 상담사이트를 링크하여 질문과 무관한 내용을 답변하였다. 이러한 형태가 블로그, 카페 글에서도 빈번하게 발견되고 이로 인해 사용자들은 질문에 부합하는 정보를 얻는데 어려움이 있다.

본 연구에서는 상기에 언급된 광고의 정의를 종합하여 상업적 내용 등으로 정보 전달의 본질을 흐리는 내용의 글을 광고 글이라고 정의한다. 그리고 이러한 광고 글에서 광고 문단을 필터링하는 시스템을 제안한다.

SNS 댓글 상에 올라오는 불법 이미지 홍보물들을 필터링하기 위한 선행연구에서는 OCR을 적용하여 영상에 삽입되어 있는 텍스트를 추출한 후, 형태소 분석기를 통해 명사를 추출하고 미리 정의된 광고 키워드와 비교하여 광고 여부를 판단한 후 필터링하는 시스템이 제안된 바 있다[8][9]. 현재 사용자들이 주로 사용하는 광고 차단 프로그램인 독일의 Eyeo사가 제공하는 Adblock의 경우, 브라우저 확장 도구로 사용자의 웹 화면에서 HTTP 요청과 아이프레임, 스크립트, 플래시를 차단하여 이미지 형태의 광고를 제거해주는 서비스를 제공한다[10]. 그러나 이러한 필터링 방법들은 학습된 모델을 통해 광고를 판별하여 차단하는 개념이 아니기 때문에 상업적 목적이 없는 이미지 형태의 공익 광고들을 차단하는 경우가 생긴다. 또한, 최근 광고의 주된 형태는 글로 작성된 바이럴 마케팅이며 기존 광고 필터링 시스템들은 최근 광고 형

태에 적용될 때 한계점이 존재한다.

3. 광고 글 필터링 모델 제안

본 연구의 필터링 시스템은 기존에 사용되던 광고 필터링 시스템들의 방식과 달리 글의 문맥을 파악하고 광고 유무를 판단할 수 있는 모델이다. 광고 유무를 판단할 수 있도록 BERT(Bidirectional Encoder Representations from Transformers) 언어 모델을 이용하여 시스템을 구현한다. 본 시스템은 사용자의 검색 결과 중 바이럴 마케팅 형태의 광고 글이 제외되도록 문서를 가시화하는 것을 목적으로 한다. 이를 통해 바이럴 마케팅 광고 글이 증가하면서 발생했던 검색에 낭비되는 시간과 광고에 의한 사용자의 피로를 감소시킬 수 있을 것이다.

광고성 글 필터링 모델의 전체적인 구조는 그림 2와 같다. 사용자는 웹 브라우저를 통해 정보 검색을 하게 된다. 웹 브라우저에서 검색하면서 발생하는 웹페이지 데이터(Web page Data)는 HTML 데이터 분석 시스템(HTML Data Analysis System)을 통해 웹페이지 태그 분석 및 문서 추출을 진행한다. 추출된 문서정보의 내용에는 광고만으로 구성된 형태가 아닌 광고와 정상 정보가 같이 포함되어있는 특징이 있다. 이를 고려하여 문단 단위로 광고판단 시스템(Automatic Advertisement

Judgment System)으로 전송되어 자연어처리(NLP: Natural Language Processing, 이하 NLP)를 진행하게 된다. 본 자연어 처리에서는 단순하게 광고에서 많이 사용되는 특정 단어의 빈도 등을 분석하는 것이 아닌 임베딩 모델을 사용하며 문단 단위로 문맥을 파악하여 높은 정확도의 문서분류(Document Classification)를 진행한다.

광고 판단 시스템으로부터 반환된 광고 유무와 문단을 레이블링(Labeling)해 구별된 문서 셋(Classified Document Set)을 웹페이지 가시화 시스템(Web page Visualization System)에 전송한다. 웹 가시화 시스템에서는 웹페이지 내 광고 판단이 완료된 문단 정보의 가시적 효과를 위해 CSS를 수정하고 필터링된 문서(Filtered Document)를 웹 페이지에 반영한다. 사용자는 불필요한 광고 내용이 블라인드 처리된 웹 페이지를 브라우저를 통해 확인할 수 있다.

광고 필터링 경로(Ad Filtering Path)를 통한 처리 과정에서 광고 분류의 성능의 한계로 오분류된 경우가 발생할 수 있다. 본 문제를 개선하기 위해 광고 필터링 개선 경로(Ad Filtering Performance Improvement Path)를 고안했다. 해당 경로에서는 False Negative(FN), False Positive(FP) case와 같이 정상적으로 필터링 되지 않은 문서를 소비자가 신고하도록 유도하였다. 사용자는 오분류된 문서와 광고 유무를 체크하여 신고하게 되면 학습 데이터 양식의 라벨(Label)에 맞게 데이터가 가공되며 이는 데이터베이스(Database)에 누적된다.

데이터베이스에 누적된 데이터 셋은 주기적으로 학습 서버(Training Server)에서의 학습을 진행하여 개선된 가중치(Improved Weight)를 광고 판단 시스템에 적용한다. 본 과정을 통해 광고 필터링 성능을 개선한다.

4. 광고 글 분류 및 적용

4.1 광고 글 분류 방안

광고 글에 들어갈 수 있는 광고의 범위는 매우 다양한 주제를 특정하기 어렵다. 정보검색과 텍스트 마이닝에서 이용하는 가중치인 TF-IDF 방식을 적용하여 단어와 문서 단위의 빈도수를 파악해 광고 유무를 파악하는 방법의 경우 대량의 빅데이터를 수집 및 가공해야 높은 성능을 기대할 수 있고 가공하더라도 분석을 통한 공통 단어 나 문단이 일관되지 않을 경우 높은 정확도를 기대하기 어렵다. 또한, 최근 광고는 바이럴 마케팅 형태를 가지는

경우가 많아서 표면적으로 광고처럼 보이지 않지만 문맥을 고려하여 자세히 읽었을 경우 광고 특성을 보이는 경우가 많아 빈도수 기반의 분석 방법은 높은 성능을 기대하기에 한계가 있다[11].

Word2Vec 또한 단어 수준의 임베딩을 수행하기 때문에 광고의 광범위한 특성을 고려하기 어렵다[12].

이러한 문제점을 고려하기 위해 문단 단위로 학습할 필요가 있으며 문단 수준의 임베딩을 수행하는 모델은 GPT, ELMo 등이 있다. 본 연구에서는 이 중 가장 높은 정확도를 보이는 BERT 언어 모델을 사용한다[13].

BERT 언어모델은 어휘의 양방향 문맥 정보와 문장 간 선후관계를 학습하여 단어와 문맥을 반영한 벡터로 표현한다[14]. 이를 본 시스템에 적용하여 광고성 글에 대한 문맥을 파악하고 분류하여 광고성 글로 판단되는 문단만을 비가시화시켜 사용자에게 필요한 내용을 효과적으로 전달한다.

Google에서 공개한 다국어 BERT는 Fine-tuning만으로도 여러 응용이 가능하고 높은 성능을 보여주었으나 한국어에 특화되지 않았다. 2019년 ETRI에서 공개한 한국어 기반 언어모델인 KorBERT는 Google의 다국어 언어모델보다 한국어 분석 단위에 특화되었고 더 좋은 성능을 보였다[15]. 이러한 성능 차이로 인해 본 시스템에는 한국어로 사전 학습된 KorBERT 언어모델을 사용한다.

4.2 광고 글 필터링 학습

2장에서 정의한 광고 글과 정상적인 정보를 포함하는 글을 분류하기 위해 표 1과 같이 데이터 셋을 구성해 학습을 진행한다.

표 1과 같이 광고 글로 판단되지 않은 문장은 Label을 0으로, 광고 글로 판단된 문장은 Label을 1로 설정해 학습을 진행한다. 광고는 주제가 다양하고 광범위하기 때문에 각 주제별로 학습하여 결과를 확인하고 최종적으로 데이터를 통합하여 학습한다.

Table 1. Examples of Training Data

Class	Label	Content
Ad X	0	정보처리기사 응시자격을 단기간에 맞출 수 있는 방법은 학점은행제 방법이 있습니다.
Ad O	1	저와 같은 1:1 맞춤 컨설팅 전문가와 정보처리기사 응시자격을 학습플랜을 계획하시고 시작하시길 꼭 추천드립니다~!

4.3 광고 비가시화

필터링 모델을 적용함에 있어 시스템적 판단뿐만 아니

라 사용자가 직접 보고 사용할 수 있는 가시화 시스템이 필요하다. 이를 위해 HTML 문서의 CSS 수정을 통해 정보와 광고의 가시적 차이를 통해 차이를 한눈에 확인할 수 있는 비가시화 시스템을 개발하였다.

검색 및 필터링, 비가시화 과정의 진행은 Selenium 브라우저를 이용해 진행한다. 정보검색 과정에서 사용자가 검색한 글에 대한 답변 내용 중 광고 여부 판단이 필요할 경우 광고 여부 판단 기능을 실행하여 답변 내용을 크롤링해 수집한다. 수집된 답변 내용 중 사용자의 정보 검색 목적에 부합하는 정보와 광고가 공존하여 실질적인 광고를 분류하기 위해서는 문단 단위 필터링 시스템 적용이 필요하다. 이를 위해 글을 문단 및 어절 단위로 분류하고 인덱스를 추가하며 분류된 결과 값에 따라 광고 글 필터링 시스템을 적용한다. 시스템 적용 후 광고라고 판단된 문단의 단어 및 부여된 인덱스를 이용해 광고성 문단에 존재하는 단어들의 위치로 접근한다.

1달 과정으로 교육이 진행되며 Python의 전체적인 부분에 대해서 공부할 수 있습니다. 뿐만 아니라 C언어, C++, 자바스크립트, JAVA 등 다양한 프로그래밍 언어에 대해서도 교육이 진행되고 있으니 많은 도움이 될것하네요.

서울, 대구, 부산에 위치한 IT전문교육기관에서 진로 및 취업 컨설팅을 진행중인 전문컨설턴트로서 기본적인 교육에 대한 안내부터 시작하여 학생들의 상황과 적성에 따라서 다양한 부분에 있어서 1:1 무료컨설팅 진행중이나 부담없이 문의주시기 바랍니다.

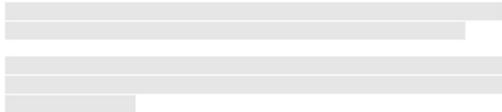


Fig. 3. Application Example through Ad Filtering Model



Fig. 4. Ad Filtering Model Interface

접근한 단어들에 대해 그림 3과 같이 태그단위 CSS 스타일 추가를 통해 광고 글을 원문과 다르게 블라인드 처리하여 정보 글과 광고 글에 차이를 둔다. 결과적으로 사용자의 검색 결과 중 실질적인 정보에만 접근할 수 있도록 가공하여 가시적인 문서를 제공한다. 이후 필터링된 내용의 확인이 필요한 경우를 대비하여 해당 글 위에 마우스 커서를 올리거나 적용된 CSS를 제거하는 원문보기 기능을 활용하여 본문 내용의 확인이 가능하다. 상기

의 기능들은 그림 4와 같이 PyQt5 모듈을 사용하여 GUI(Graphical User Interface)로 구현하였다.

4.4 광고 필터링 향상 방안

현재 사용되는 광고 필터링 시스템들은 웹페이지 상에서 사용자에게 자율적으로 광고에 해당하는 태그를 차단할 수 있는 기능과 광고 영상에 해당하는 링크를 신고할 수 있도록 지원하고 있다. 이를 통해 필터링 시스템은 성능의 한계를 보완한다.

본 연구의 필터링 시스템에서도 신고 기능을 지원하여 광고 필터링 기술을 보완할 수 있도록 설계하였다. 사용자가 광고 판단 기능을 사용하였을 시, 학습 모델이 완성하지 못하여 광고를 제대로 처리하지 못하는 경우가 생길 수 있다. 이런 경우 사용자 또는 관리자가 해당 문단을 입력하여 광고유무를 선택 후 신고할 수 있도록 하였다. 신고 데이터는 데이터베이스에 축적되도록 하고 주기적으로 데이터베이스의 신고 데이터를 통해 모델의 재학습과정을 거쳐 업데이트하는 방식으로 보완시켜 필터링 시스템의 성능향상을 돕는다.

5. 실험 및 결과

실제 사용자가 경험하는 검색과 가장 유사한 실험환경을 구성하기 위해 2019년 기준 94.7%가 사용하는 네이버[16] 지식iN을 기준으로 실험을 진행하였다. 광고는 다양한 분야에 적용되어 광범위하다는 특징을 가지고 있다. 이 특징에 맞춰 각 주제별로 데이터셋을 구성해 성능을 평가하고 데이터를 확장하여 적용하였다.

Table 2. Composition of Training Dataset

Subject	Ad	Training	Evaluation
Language	X	200	200
	O	144	143
Job Seeking	X	300	300
	O	171	171
Integration	X	800	800
	O	496	497

학습 데이터 셋과 평가 데이터 셋의 구성은 표 2에 나타난 바와 같으며 통합 데이터의 경우 언어, 구직 외 3개를 포함해 총 5개의 주제로 이루어져 있다.

Table 3. Performance Evaluation of Ad Filtering Model

Subject	Accuracy	Precision	Recall	F1 Score
Language	91.8	90.8	89.5	90.1
Job Seeking	90.4	85	89.5	87.2
Integration	89.2	84.3	87.6	86

각각의 데이터 셋에 대한 결과는 표 3과 같으며 각 주제별 결과보다 통합 데이터의 결과가 비교적 낮게 나왔지만 광범위한 비정형적 특징을 가진 광고의 특성에도 불구하고 높은 성능을 보였다. 광고글 분류를 위한 주제가 늘어남에도 정확도를 포함한 여러 분류성능 지표가 높은 수치를 유지하는 결과를 보여준다. 이에 제안된 광고글 필터링 모델을 적용될 경우 지속적인 성능 향상이 기대된다.

6. 결론

최근 인터넷 사용자의 증가와 스마트폰 등 인터넷을 사용할 수 있는 기기들이 증가하면서 대부분의 사람들이 일상적으로 인터넷을 사용하고 있다. 인터넷 사용량이 늘어남에 따라 인터넷 정보량 또한 기하급수적으로 늘어나고 있으며 기업들은 클라우드, SNS 등 다양한 서비스를 제공하고 자신들의 제품과 서비스를 인터넷을 통해 홍보하고 있다.

그러나 최근 상업적인 목적으로 무분별하게 광고 글을 게시해 인터넷 정보를 흐리는 바이럴 마케팅이 발생하고 있다. 때문에 정보검색을 목적으로 하는 소비자의 정보 습득 시간을 증가시키고 편의성을 감소시키는 문제점이 대두되고 있다. 이러한 문제점을 해결하기 위해 AdBlock, OCR을 사용하는 연구 등의 방법이 사용되었으나 대부분 글로 작성된 인터넷 정보들 속에서 무분별한 광고를 제거하는 데 한계를 가졌다.

본 논문에서는 이러한 한계점을 극복하기 위해 KorBERT를 사용하여 글의 앞뒤 문맥을 파악하고 광고의 성격을 가진 글만 블라인드 처리해 정보검색을 하는 사용자의 편의성을 증가시키는 방안에 대해 제안하였다. 사용자들이 정보검색을 위해 사용하는 질의응답 서비스인 네이버 지식iN을 기반으로 답변 내용을 크롤링해 데이터를 수집하고 광고 유무에 따라 라벨링하여 학습을 진행하였다. 이에 대한 결과를 검증하기 위해 실험을 진

행하였으며 5개의 주제를 통합한 데이터에 대한 정확도(Accuracy), 정밀도(Precision)는 각각 89.2%, 84.3%의 결과를 나타냈고 광고의 비정형적 특성으로 인한 어려움에도 불구하고 높은 성능을 보였다. FN, FP와 같이 오분류된 결과를 보이는 경우 사용자에게 치명적인 결과를 초래하진 않으나 신고 기능을 통해 지속적으로 성능 개선이 되도록 하였다. 이를 통해 사용자는 무분별한 광고 내용 속에서 정보 습득 시간을 줄일 수 있을 것으로 기대된다.

추후 연구를 통해 네이버 지식iN 서비스를 넘어 다양한 환경으로부터 획득한 데이터를 수집하고, 광고 필터링에 관한 최신 연구결과와의 비교실험 및 성능평가를 통해 제안하는 모델의 성능을 개선할 예정이다.

References

- [1] ITU(International Telecommunication Union), Individuals using the Internet, 2005-2019 [Internet], Available From: <https://www.itu.int/en/ITU-D/Statistics/Pages/stat/default.aspx> (accessed Jun 25, 2020)
- [2] David Reinsel, John Gantz, John Rydning, The Digitization of the World From Edge to Core, IDC(International Data Corporation), pp.1-28, Nov. 2018.
- [3] Jeil Oh, "Supreme Court of Korea 'Portal screen change software release, No business interruption' " [Internet], NEWSIS, cited May. 05 2016, Available From: <https://bit.ly/2VcIZWR> (accessed Jun. 23, 2020)
- [4] JeongHoon Lee, "Online Advertising by Search Engine and Criminal Responsibility", Hongik Law Review, Vol.19, No.4, pp.59-81, 2018. DOI: <http://dx.doi.org/10.16960/jhrl.19.1.201802.59>
- [5] Byunghee Kim, "New Definitions and Ranges on Advertising : Mixed Methods Approach", KJA(The Korean Journal of Advertising), Vol.24, No.2, pp.225-254, 2013.
- [6] Korea Industrial Marketing Institute, "Viral Marketing", Marketing, Vol.49, No.5, pp.59-67, May. 2015.
- [7] Anonymous. Driver's insurance cost in the 20s [Internet]. NAVER, [cited Jun 13 2020], Available From: <https://bit.ly/2OaM31W> (accessed Jun. 24, 2020)
- [8] Hyoung-Woong Yoon, Namsu Kim, Sohyun Park, Ye-Bin Jeong, Hae-Yeoun Lee, "SNS Advertisement Filtering System", Proceedings of KIIT Conference, pp.261-262, Jun. 2018.
- [9] Ji-A. Kim, Geum-Boon. Lee, "An Effective Method for Blocking Illegal Sports Gambling Ads on Social Media", Journal of the Korea Society of Computer and

Information, Vol.24, No.12, pp.201-207, Dec. 2019.
DOI : <https://doi.org/10.9708/iksci.2019.24.12.201>

- [10] Taewon Song, Heeok Lee, "A Study of Legal Aspect on the Blocking Ads by Adblock", InHa Law Review, Vol.19, No.2, pp.147-175, 2016.
- [11] Sunju Park, Seungwha Chung, Naseong Pyo, Soonki Hwang, "Bottlenecks in Building an Online Customer Base: A Experimental Field Study on Viral Marketing", The Journal of the Korea Contents Association, Vol.19, No.1, pp.682-695, 2019.
DOI : <http://dx.doi.org/10.5392/JKCA.2019.19.01.682>
- [12] Sangheum Hwang, Dohyun Kim, "BERT-based Classification Model for Korean Documents", The Journal of Society for e-Business Studies, Vol.25, No.1, pp.203-214, Feb. 2020.
DOI : <http://dx.doi.org/10.7838/jsebs.2020.25.1.203>
- [13] Soojong Lim, Hyunki Kim, "Current Status of Deep Learning Pre-training for Natural Language Processing and Application to Korean" Institute of Culture Convergence Archiving, Vol.2, No.2, pp.111-118, Oct. 2019
- [14] Devlin, Jacob, et al. "Bert: Pre-training of deep bidirectional transformers for language understanding" Association for Computational Linguistics, pp.4171-4186, Jun. 2019.
DOI : <http://dx.doi.org/10.18653/v1/N19-1423>
- [15] ETRI, Results of KorBERT language model comparison with Google [Internet], ETRI, Available From: http://aiopen.etri.re.kr/service_dataset.php (accessed Jun. 24, 2020)
- [16] DMC MEDIA, 2019 Portal Site Usage Behavior Survey Analysis Report - Summary, Analysis Report, DMC REPORT, Korea, pp.1-13

박 래 근(Raegun Park)

[준회원]



• 2015년 3월 ~ 현재 : 상명대학교
공과대학 스마트정보통신공학과
재학 중.

<관심분야>

인공지능, 자율주행, 컴퓨터 비전, 자연어처리

윤 혁 진(Hyeok-Jin Yun)

[준회원]



• 2015년 3월 ~ 현재 : 상명대학교
공과대학 스마트정보통신공학과
재학 중.

<관심분야>

인공지능, 자율주행, 컴퓨터 비전, 자연어처리

신 의 철(Ui-Cheol Shin)

[준회원]



• 2015년 3월 ~ 현재 : 상명대학교
공과대학 스마트정보통신공학과
재학 중.

<관심분야>

인공지능, 안드로이드 어플리케이션 개발, 프론트엔드

안 영 진(Young-Jin Ahn)

[준회원]



• 2015년 3월 ~ 현재 : 상명대학교
공과대학 스마트정보통신공학과
재학 중.

<관심분야>

인공지능, 빅데이터, 프론트엔드

정 승 도(Seungdo Jeong)

[중신회원]



- 2001년 2월 : 한양대학교 전자통신전파공학과 (공학석사)
- 2007년 8월 : 한양대학교 전자통신전파공학과 (공학박사)
- 2012년 3월 ~ 2015년 2월 : 한양사이버대학교 조교수
- 2015년 3월 ~ 현재 : 상명대학교 스마트정보통신공학과 조교수

〈관심분야〉

멀티미디어 색인, 정보검색, 패턴인식, Tensor 응용